IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR LETTERS PATENT

BY

Paul Kirkby

42 St. Johns Avenue
Harlow
Essex
UNITED KINGDOM

FOR

TRAFFIC FLOW MANAGEMENT IN A COMMUNICATIONS NETWORK

12599ID Kirkby

TRAFFIC FLOW MANAGEMENT IN A COMMUNICATIONS NETWORK

FIELD OF THE INVENTION

This invention relates to methods and apparatus for controlling traffic flow in a communications network so as to reduce congestion.

BACKGROUND OF THE INVENTION

Increasing volumes of communications traffic are now being carried on packet networks, and in particular on Internet Protocol (IP) networks. Such networks comprise on any nodes or routers interconnected by links so as to define a mesh. A recent introduction has been the concept of network having an optical core in which traffic is carried on switched optical fibre paths between routers. The core is accessed by an edge network Typically in the design of high capacity IP networks, routers are classified as either core routers or edge routers. Edge routers carry out all the network ingress and egress functions in particular controlling the incoming traffic streams across the network. Core routers act as transit routers forwarding network traffic from one network node to another.

In such a network, the user data is assembled into packets and each provided with a header identifying the destination of the packet and optionally, including routing information. The header may further contain information relating to the router of the packet contents and identifying a priority class for the packet. For example, packets containing high quality of service real time traffic, and is voice, will be accorded the highest priority, while packets containing 'best efforts' data may be accorded a low priority.

A particular problem that has been experienced with certain types of traffic, particularly data traffic and real-time video traffic, In its inherently bursty nature. Further, this burstiness occurs on a timescale that is shorter than feasible network control loop timescales, and thus can lead to congestion when traffic is heavy.

When congestion occurs, ordinary data traffic which is not critcally time sensitive can be briefly buffered in the routers which are experiencing congestion. Urgent data traffic and real time interactive services such as voice and video cannot be delayed.

In order to maximise the overall network utilisation, it is desirable to perform statistical multiplexing of traffic traversing the network while providing a prior allocation of resources and protection particularly for the delay sensitive traffic. Existing control and feedback mechanisms Such as described in patents...are however inadequate to respond to this bursty traffic at a sufficiently rapid rateto provide this resource allocation and protection. In the conventional approach to this problem, the high speed statistical variations in traffic flow are simply allowed for by setting large margins in the setting of control levels for determining feedback price. Proposals for 'pricing' ingress flows at the edge of the network for admission control purposes have involved for instance measuring the 'effective bandwidth' of the flow. (Ref) Effective bandwidth is a measure of the bandwidth that needs to be reserved to give a desired packet loss ofr delay rate on a statistically varying flow. Unfortunately effective bandwidths do not add linearly on aggregation so are difficult to use in a congestion price feedback control scheme.

SUMMARY OF THE INVENTION

An object of the invention is to minimize or to overcome these disadvantages.

According to a first aspect of the invention there is provided a method of controlling traffic flow in a communications packet network, the method comprising determining for flows within the network a mean utilisation requirement and a measure of a variance from that mean, and determining from said mean and variance and bandwidth pricing so as to control the admission of said flows to the network.

Advantageously, control limits are set by the network operator for mean and variance, and the pricing is increased as one or both of these imits is approached so as to provide feedback control of admission to the network.

The control method may be embodied as software in machine readable form on a storage medium.

According to another aspect of the invention, there is provided a method of controlling admission of traffic flows to a communications network, the method comprising sampling the traffic flows each at an ingress, and sampling an aggregate flow of said flows at some or all of the resources used by the aggregate flow, determining from said sampling a mean bandwidth requirement for each traffic flow and a measure of the variance from that mean, determining from said mean and variance measurements first and second prices for the mean and variance components of the controlled traffic flows that are admitted to the network, and determining from said first and second prices an admission cost for each said flow so as to regulate the admission of that flow.

According to a further aspect of the invention, there is provided an admission control arrangement for a communications network, the arrangement comprising sampling means for sampling a traffic flow, means for determining from said sampling means a measure of mean bandwidth requirement and of a variance from that mean, and price computation means for determining from said mean and variance a cost or price for bandwidth case so as to provide ingress price control for admission of the traffic flow to the network.

The admission control arrangement may be incorporated in a network manager, e.g. in software form.

In our arrangement and method, overall resource utilisation is optimised by the use of congestion price based feedback that separates out the components of price that affect the mean traffic flow from the standard deviation of the flow. It is well known that on aggregation, standard deviations add according to a square route law and means add linearly. In a mixed traffic flow type network, a network flow optimisation process that separates these components can provide improved overall resource utilisation. A further advantage of this arrangement and method is that that it enables a fairer allocation of 'network –price' to any given ingress flow. Rather than simply allocating the network charges on the basis of mean ingress flow, the charges are based on a measure of mean and standard deviation. This

discourages users from abusing the network by sending excessively bursty traffic, unless it is genuinely important to the user. It should be noted that the terms prices and charges as employed herein refer to an internal network currency used for admission control purposes and is generally separate from any real billing.

BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments of the invention will now be described with reference to the accompanying drawings in which:-

Figure 1 is a selective design of an exemplary network;

Figure 2 illustrates the aggregation of traffic flows at a node in the network of Figure 1;

Figure 2a shows the construction of an ingress controller;

Figure 3a and 3b illustrate idealised and practical bandwidth demands for a traffic flow.

Figure 4 illustrates the variance of the traffic flow of Figure 3;

Figure 5a illustrates the measured mean and standard deviation of a traffic aggregate; and

Figure 5b illustrates the primary flow control level and its division into separate allocations for variance and mean..

DESCRIPTION OF PREFERRED EMBODIMENTS

Referring first to Figure 1, this shows an exemplary network in schematic form. As shown in Figure 1, the network comprises a number of nodes 11 interconnected by links 12. As shown in Figure 1, the network comprises a core region 13 accessed via an edge region 14. The links 12 are usually optical fibre, particularly in the core region. Advantageously the network of Figure 1 is an Internet Protocol (IP) or MPLS (multi protocol label switched) network in which traffic is transported in packet form. An objective of the resource control system is to control the ingress of IP and MPLS traffic in such a way that different traffic classes are treated optimally. In particular, delay sensitive classes of traffic must see minimal congestion inside any router or on entering any of the packet buffers at the entrance to each optical link.

Referring now to Figure 2, this shows in schematic form a an edge node 11 and a core network node 11a. At the edge node 11, a number of input traffic flows are aggregated on to a traffic path on link 12. At the core node 11a two transit traffic flows are shown merging. Prior to entering the packet buffer 17 at the entrance to transmission link 12, the traffic is sampled by the aggregate sampling circuit 21. At the entrance to the next link on the traffic path 12a a similar measurement arrangement exists 17a,21a. The traffic flows are sampled by sampling circuits 21 and 22 to determine both the mean bandwidth usage x_i and the standard deviation σ_i from that mean for each aggregated flow across the network. The mean and standard deviation measurement are processed by a network admission controller to determine a pair of prices for using that particular resource. This price pair defines separate prices for the mean component of traffic flow and the deviation (or variance) component. A separate ingress controller 23 (Figure 2a) in the edge router samples 24 and measures the mean and deviation of individual edge to edge flows entering the network. The ingress controller also continuously monitors the sum of the resource price pairs for the edge to edge path it is using. (note there is one ingress controller per edge to edge path, the explicit edge to edge path being defined for instance by MPLS labels attached to each packet. The user (or a software object using pre-agreed ingress control rules) can then either accept this price or modify this mean bandwidth or standard deviation bandwidth requirements to obtain his optimum quality of service vs price. To modify the ingress traffic flow the ingress controller could use the traffic shaper 25 or for example signal back to the original traffic source (not shown). The traffic shaper controls scheduler 26. This price feedback mechanism provides a self-regulating mechanism on the bandwidth demands imposed on the network.

The ingress control 23 controls traffic on every end to end path through the network. The paths may be MPLS paths.

Referring now to Figures 3a and 3b, these illustrate respectively idealised and practical bandwidth requirements for a traffic flow. Figure 3a shows a slowly varying mean bandwidth demand, while figure 3b shows a typical rapid short term variation superimposed on the mean.

The short term variations, which represent the deviations of the traffic flow from the mean, are illustrated in Figure 4. These can be statistically analysed in real time to

give a measure of mean and standard deviation. This analysis also gives the variance which is the standard deviation squared. A variety of algorithms could be used for this purpose. The exponentially weighted time averaged mean is one of the simplest, whilst the sampled data can and the mean together can be used to give the time averaged variance. The rest of this description makes the assumption that the sampling periods and rate and time averaging parameters used for ingress control purposes are sufficiently similar to those used by the aggregate flow meters that insignificant errors are caused. Typically the time averaging time constant is greater than the feedback time constants of the DRC control system. Typically 100ms to 10 seconds. The variance sampling period is typically sufficiently short that the shortest bursts you are interested in controlling are captured. This implies a sampling time period does not need to be much shorter than the specified worst case permitted delay variation for the traffic being controlled. 1-10ms might be appropriate for typical delay sensitive traffic. (Longer sampling periods could be tolerated if margins were increased. A measure of the significance of these rapid short term variations is given by the standard deviation (σ) or by the corresponding variance σ^2.

In the arrangement of Figure 2a in which a number flows are aggregated on a common path. Assuming that traffic deviations are uncorrelated in time, the standard deviation σ_A of the aggregate flow is given by the expression

$$\sigma_A = \sqrt{\sum \sigma_i^2}$$

The aggregated mean traffic of course adds linearly so the aggregated mean flow is given by the expression

$$x_A = \sum x_i$$

Referring now to Figure 5a, this illustrates a bandwidth plan from which pricing information is determined to provide feedback control for admission to the network. In figure 5a, the network operator sets a peak bandwidth maximum or control level x_c which, in the ordinary cause of events should not be exceeded, even momentarily by the peak bursts of the aggregated traffic. A typical flow has a mean bandwidth x_A well below this maximum level. Further, to minimise the risk of

congestion, this mean x_A should be at least one and preferably 'k' standard deviations below this control level x_C. To ensure that the probability of momentary congestion is sufficiently small for all practical purposes k should typically lie in the range 3 to 6.

In a preferred embodiment, separate price calculations are performed from the mean x_A and standard deviation σ_A demands on the network as follows

The network operator subdivides the control level bandwidth into two sub allocations, x_{CM} the allocation for the mean flow, and x_{CD} the nominal allocation for the components of flow that deviate from the mean. These two allocations are related to the control level by the equation $x_C = x_{CM} + x_{CD}$,as illustrated in Figure 5b. . We determine the mean pricing P_x as a function of the mean demand and the control level for mean traffic. This price will typically be quoted by the resource in terms of a price per unit bandwidth.

$$\text{i.e. } P_x = F\ (x_{CM}, x_A)$$

In a discrete system, in which price is updated at regular intervals , Δt, an example function might be.

$$P_x(t+1) = |P_x(t) + \beta\Delta t\,(x_{CM} - x_A)|_+$$

This equation indicates that the new price at the end of the current time interval equals the current price plus a gain factor times the time interval and the difference between the control level for mean traffic and the current mean flow. The outer brackets with subscript + sign signifies that the solution is constrained to be positive. The parameter β is the feedback gain. This is an example of an integrating control feedback system, as the error signal is integrated over time to produce change in price. Other variants include differential and proportional feedback.

Similarly, we define the variance pricing P_V in terms of the the measured standard deviation σ_A and the control level x_{CD}. This price will typically have units of price per unit time per unit bandwidth variance. i.e..

$$P_V = F\ (x_{CD}, \sigma_A)$$

In a discrete system an example function might be

$$P_V(t+1) = |P_V(t) + \alpha\Delta t\,(x^2_{CD} - k^2\sigma^2_A)|_+$$

This equation indicates that the new variance price at the end of the current time interval equals the current price plus a gain factor times the time interval and the difference between the square of control level for deviation and the square of k times the measured aggregate variance. The outer brackets with subscript + sign signifies that the solution is constrained to be positive. The parameter α is the feedback gain. This is an example of an integrating control feedback system, as the error signal is integrated over time to produce change in price. Other variants include differential and proportional feedback.

In both these examples the price increases rapidly as the argument of the inner bracket (the error signal) goes positive.

A more sophisticated pricing may be obtained by introducing a second order term which takes into account the first and/or second differentials of the traffic flow means and deviations. . In that case the pricing functions become

$$P_m = F\left((x_{CM}, x_A), \frac{dx_A}{dt}, \frac{d^2x_A}{dt^2}\right) \text{ and } P_V = F\left((x_{CD}, \sigma_A, \frac{d\sigma_A}{dt}, \frac{d^2\sigma_A}{dt^2}\right)$$

The price that is fed back to the user is in the form of the two separate prices P_m and P_V. These prices can be quoted in terms of a price per unit bandwidth and a price per unit variance. The total 'price' charged to the user is then determined by measurements on the traffic leaving the user of the mean and variance of his traffic. These are multiplied by the prices per unit and added together to give the total price to the user T. This enables the network (and optionally) the user to see a much fairer estimate of the network 'cost' to transport that particular flow. This information can then be used for admission control purposes, for instance for smoothing the flow to reduce the variance component of the flow or reducing or even shutting off the mean flow if the total price T is becoming excessive. It can also be used for admission control of inelastic traffic flows where a user would ask for admission of a guaranteed flow with certain quoted mean and variance characteristics. Once admitted, the ingress controller of such inelastic traffic flows

is committed to allow the ingress of traffic within the pre agreed limits, irrespective of how high the internal congestion price subsequently goes.

The technique enables finely differentiated services to be provided by control of ingress flows at the network edge. Some users may, for instance want a class of service which highly values the transmission of bursty traffic without delay. These users would have a service level agreement that assigned a high WtP (willingness to pay) to the variance component of the traffic. Thus even in high congestion the bursts from such users would be allowed into the network whilst other users with a low WtP for variance would have their traffic bursts smoothed and hence momentarily delayed.

A modification of the schemes described above is a system in which the sub-allocations of bandwidth for mean and standard deviation control levels x_m and x_{CM} are not defined by management fiat, but are estimated dynamically from the measurements of mean and deviation parameters of the incoming aggregate flows. The management system still defines the bandwidth peak maximum control level x_C. The simplest way to determine the standard deviation allocation is simply to define

$$x_{CD} = x_C - x_A$$

That is to say the allocation for standard deviation is simply taken to be the difference between the current mean traffic level and the peak maximum control level. This has the advantage that as the mean level rises towards the peak bandwidth control level, the variance price rises in an inverse square manner, drastically curtailing new traffic bursts. When the network has only low mean traffic, the variance price is extremely low and bursty traffic is hardly discouraged at all.

In this case it would be natural to define the control level for the mean flow as

$$x_{CM} = x_C - k.\sigma_A$$

That is to say the mean control level is defined as the peak bandwidth control level minus K times the currently measured standard deviation. In practice caution may suggest that other constraints such as a minimum bandwidth margin between the peak control level and the mean control level are added.

For optimum resource usage in a future system in which the ingress controllers actively adjusted the shaping of the mean and deviation components of their traffic in response to mean and deviation prices and the class of traffic being transmitted, then maximum user utility would be obtained when proportional fairness is applied to the allocation split between mean and deviation traffic. The resource would thus adjust the ratio of x_{CM} to x_{CD} to be the same as the current revenue from mean traffic and the current revenue the variance component of the traffic. These adjustments would have to be carried out slowly in comparison to the DRC feedback control time constant or instability could result.

The price for traffic traversing a series of resources along a path is optimally found by adding the prices at each resource linearly. Thus, the prices per unit bandwidth for the mean flows for traffic flowing along a particular path are found by adding the prices per unit bandwidth of the means in each resource that the path utilizes. We have found that, for optimum results, the price per unit variance should be determined by adding the price per unit variance of each resource that the path utilizes, rather than adding the price per unit standard deviation.

In a further modification, particularly bursty traffic can incur a further price penalty by the user of an exponential weighting on the standard deviation measurement. This exponential weighting function can be used to weight large positive deviations of traffic flow from the mean more strongly than small deviations. This would be used as a second order correction to discourage users from transmitting large bursts of traffic unexpectedly at very high rates. Note that the use of variance rather than simple deviation from a mean already penalizes large sudden deviations with a square law dependency.

It will be understood that the above description of a preferred embodiment is given by way of example only and that various modifications may be made by those skilled in the art without departing from the spirit and scope of the invention.